IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION FOR LETTERS PATENT

A Contrast-Based Image Attention Analysis Framework

Inventors:
Yu-Fei Ma
Hong-Jiang Zhang

TECHNICAL FIELD

[0001] The invention pertains to image analysis.

BACKGROUND

[0002] Effective information retrieval from a large image library is generally a function of subject matter retrieval accuracy and an adaptive image display scheme for suitable presentation by a variety of different and computing devices, many of which are of compact design. Such small form factor computing devices include, for example, handheld computing and/or communication devices, many of which have limited display, processing, and/or memory capabilities. The question of how to identify important/representative regions of an image is related to both retrieval accuracy and adaptive image display. If semantics of each region of an image are known beforehand, these questions are easily solved. However, programmatic determination of image semantics is generally considered to be a machine intelligence issue and computationally intensive task, not to mention, one that is beyond capabilities of most conventional computer vision systems. Accordingly, alternatives to semantic understanding of image content are desired for improved information retrieval and adaptive image display.

SUMMARY

[0003] Systems and methods for contrast-based image attention analysis are described. In one aspect, image attention is modeled by preprocessing an image to generate a quantized set of image blocks. A contrast-based saliency map for

modeling one-to-three levels of image attention is then generated from the quantized image blocks.

BRIEF DESCRIPTION OF THE DRAWINGS

[0004] In the figures, the left-most digit of a component reference number identifies the particular figure in which the component first appears.

[0005] Fig. 1 shows an exemplary computing environment within which systems and methods for generating a contrast-based saliency map for image attention analysis may be implemented.

[0006] Fig. 2 shows further exemplary aspects of system memory of Fig. 1, including application programs and program data for generating a contrast-based saliency map for image attention analysis.

[0007] Figs. 3-8 show respective images that illustrate how contrast underlies color, texture, and shape perception. In particular, figure pairs 3 and 4, 5 and 6, and 7 and 8 represent synthesized image pairs.

[0008] Fig. 9 shows an example of an original image prior to preprocessing and quantization.

[0009] Fig. 10 shows an example of the original image of Fig. 9 after quantization operations.

[0010] Fig. 11 shows an example of a contrast-based saliency map derived from the quantized image of Fig. 10.

[0011] Fig. 12 illustrates an exemplary fuzzy 2-partition of contrast-based saliency map. In particular, three layers are shown which denote a gray level higher than a (highest), s (middle), and u (lowest) layer.

[0012] Fig. 13 shows exemplary results of fuzzy growing by treating a contrast-based saliency map as a fuzzy event in view of mutually exclusive attended and non-attended areas.

[0013] Fig. 14 shows exemplary attended points derived from a contrast-based saliency map.

[0014] Fig 15-44 show respective examples of contrast-based saliency maps, attended view, attended areas, and attended points, each of which has been derived from respective original images, as illustrated.

[0015] Fig. 45 shows an exemplary procedure for generating a contrast-based saliency map for image attention analysis bases on fuzzy growing.

DETAILED DESCRIPTION

Overview

[0016] Systems and methods for generating a contrast-based saliency map for image attention analysis are described. In particular, the following framework maps contrast-based saliency via local (i.e., regional) contrast analysis. Fuzzy growing is then used to simulate human perception and to extract attended objects from the contrast-based saliency map. These attended objects include, for example, attended views, attended areas and attended points, each of which are then utilized to provide three-levels of attention data for image analysis. An attended view can effectively accelerate feature extraction during image retrieval by extracting a sub-image with information that has been objectively determined to be most important. Attended areas provide more details about substantially important areas for region-based image retrieval operations. Also, both attended

view and attended areas may facilitate quick browsing of important parts of image in a variety of display screens in different sizes. Moreover, although the attended points lack of semantics, they provide users possible search paths on images, which can be utilized to determine the browsing sequence of image regions.

[0017] In these manners, the systems and methods of the invention provide a considerably robust alternative to semantic understanding for image retrieval and adaptive image display—especially since the vision system automatically extracts attentions in images via simulation of human perception. Exemplary systems and methods for generating a contrast-based saliency map for image attention analysis are now described in greater detail.

Exemplary Operating Environment

[0018] Turning to the drawings, wherein like reference numerals refer to like elements, the invention is illustrated as being implemented in a suitable computing environment. Although not required, the invention is described in the general context of computer-executable instructions, such as program modules, being executed by a personal computer. Program modules generally include routines, programs, objects, components, data structures, etc., that perform particular tasks or implement particular abstract data types.

[0019] Fig. 1 illustrates an example of a suitable computing environment 120 on which the subsequently described systems, apparatuses and methods for generating a contrast-based saliency map for image attention analysis may be implemented. Exemplary computing environment 120 is only one example of a suitable computing environment and is not intended to suggest any

limitation as to the scope of use or functionality of systems and methods the described herein. Neither should computing environment 120 be interpreted as having any dependency or requirement relating to any one or combination of components illustrated in computing environment 120.

[0020] The methods and systems described herein are operational with numerous other general purpose or special purpose computing system environments or configurations. Examples of well-known computing systems, environments, and/or configurations that may be suitable include, but are not limited to, including small form factor (e.g., hand-held, mobile, etc.) computing devices (e.g., mobile phones, personal digital assistants—PDAs, etc.), multi-processor systems, microprocessor based or programmable consumer electronics, network PCs, minicomputers, mainframe computers, and/or so on. The invention is also practiced in distributed computing environments where tasks are performed by remote processing devices that are linked through a communications network. In a distributed computing environment, program modules may be located in both local and remote memory storage devices.

[0021] As shown in Fig. 1, computing environment 120 includes a general-purpose computing device in the form of a computer 130. The components of computer 130 may include one or more processors or processing units 132, a system memory 134, and a bus 136 that couples various system components including system memory 134 to processor 132. Bus 136 represents one or more of any of several types of bus structures, including a memory bus or memory controller, a peripheral bus, an accelerated graphics port, and a processor or local bus using any of a variety of bus architectures. By way of example, and not

limitation, such bus architectures include Industry Standard Architecture (ISA) bus, Micro Channel Architecture (MCA) bus, Enhanced ISA (EISA) bus, Video Electronics Standards Association (VESA) local bus, and Peripheral Component Interconnects (PCI) bus also known as Mezzanine bus.

[0022] Computer 130 typically includes a variety of computer readable media. Such media may be any available media that is accessible by computer 130, and it includes both volatile and non-volatile media, removable and non-removable media. System memory 134 includes computer readable media in the form of volatile memory, such as random access memory (RAM) 138, and/or non-volatile memory, such as read only memory (ROM) 140. A basic input/output system (BIOS) 142, containing the basic routines that help to transfer information between elements within computer 130, such as during start-up, is stored in ROM 140. RAM 138 typically contains data and/or program modules that are immediately accessible to and/or presently being operated on by processor 132.

[0023] Computer 130 may further include other removable/non-removable, volatile/non-volatile computer storage media. For example, a hard disk drive 144 may be used for reading from and writing to a non-removable, non-volatile magnetic media (not shown), a magnetic disk drive 146 for reading from and writing to a removable, non-volatile magnetic disk 148 (e.g., a "floppy disk"), and an optical disk drive 150 for reading from or writing to a removable, non-volatile optical disk 152 such as a CD-ROM/R/RW, DVD-ROM/R/RW/+R/RAM or other optical media. Hard disk drive 144, magnetic disk drive 146 and optical disk drive 150 are each connected to bus 136 by one or more interfaces 154.

[0024] The drives and associated computer-readable media provide nonvolatile storage of computer readable instructions, data structures, program modules, and other data for computer 130. Although the exemplary environment described herein employs a hard disk, a removable magnetic disk 148 and a removable optical disk 152, it should be appreciated by those skilled in the art that other types of computer readable media which can store data that is accessible by a computer, such as magnetic cassettes, flash memory cards, digital video disks, random access memories (RAMs), read only memories (ROM), and the like, may also be used in the exemplary operating environment.

[0025] A number of program modules may be stored on the hard disk, magnetic disk 148, optical disk 152, ROM 140, or RAM 138, including, e.g., an operating system 158, one or more application programs 160, other program modules 162, and program data 164.

[0026] A user may provide commands and information into computer 130 through input devices such as keyboard 166 and pointing device 168 (such as a "mouse"). Other input devices (not shown) may include a microphone, joystick, game pad, satellite dish, serial port, scanner, digital camera, etc. These and other input devices are connected to the processing unit 132 through a user input interface 170 that is coupled to bus 136, but may be connected by other interface and bus structures, such as a parallel port, game port, or a universal serial bus (USB).

[0027] A monitor 172 or other type of display device is also connected to bus 136 via an interface, such as a video adapter 174. In addition to monitor 172, personal computers typically include other peripheral output devices (not shown),

such as speakers and printers, which may be connected through output peripheral interface 175.

[0028] Computer 130 may operate in a networked environment using logical connections to one or more remote computers, such as a remote computer 182. Remote computer 182 may include some or all of the elements and features described herein relative to computer 130. Logical connections include, for example, a local area network (LAN) 177 and a general wide area network (WAN) 179. Such networking environments are commonplace in offices, enterprise-wide computer networks, intranets, and the Internet.

[0029] When used in a LAN networking environment, computer 130 is connected to LAN 177 via network interface or adapter 186. When used in a WAN networking environment, the computer typically includes a modem 178 or other means for establishing communications over WAN 179. Modem 178, which may be internal or external, may be connected to system bus 136 via the user input interface 170 or other appropriate mechanism.

[0030] Depicted in Fig. 1, is a specific implementation of a WAN via the Internet. Here, computer 130 employs modem 178 to establish communications with at least one remote computer 182 via the Internet 180. In this example, the remote computer 182 happens to be a small form factor device in the embodiment of a mobile telephone with a small display screen. The remote computer is representative of all possible types of computing devices that can be coupled to the computer 130 as described.

[0031] In a networked environment, program modules depicted relative to computer 130, or portions thereof, may be stored in a remote memory storage

device. Thus, e.g., as depicted in Fig. 1, remote application programs 189 may reside on a memory device of remote computer 182. The network connections shown and described are exemplary. Thus, other means of establishing a communications link between the computing devices may be used.

Exemplary Application Programs and Data

[0032] Fig. 2 is a block diagram that shows further exemplary aspects of system memory 134 of Fig. 1, including application programs 160 and program data 164 for generating a contrast-based saliency map for image attention analysis. In this implementation, application programs 160 include, for example preprocessing module 202, contrast computation and normalization module 204, attended point extraction module 206, attended area extraction module 208, and attended view extraction module 210. Aspects of these computer-program modules and their operations are now described in detail in reference to exemplary images of Figs. 3 through 44.

Contrast-Based Saliency

[0033] Contrast is an important parameter in assessing vision. Clinical visual acuity measurements generally rely on high contrast, that is, black letters on a white background. However, objects and their surroundings are typically of varying contrast. Therefore, the relationship between visual acuity and contrast allows a more detailed understanding of human visual perception. Traditional image processing techniques usually consider an image by three basic properties, color, texture, and shape. Although these techniques have been successfully applied to a number of applications, they cannot provide high level understanding

of an image, because humans usually do not perceive images from color, texture, and shape aspects separately. The systems and methods of the invention address these limitations of conventional systems by utilizing contrast attention analysis. Contrast attention analysis is especially pertinent to image analysis. Whether an object can be perceived depends on the distinctiveness (i.e., contrast) between the object and its environment. Moreover, contrast perception underlies each of the separate components of color, texture, and shape perception.

[0034] Figs. 3-8 illustrate how contrast underlies color, texture, and shape perception. For purposes of discussion, aspects of these figures are described in terms of color other than the various shades of grayscale color that are shown in the figures. In particular, figure pairs 3 and 4, 5 and 6, and 7 and 8, each represent respective pairs of synthesized images. In Figure 3, there is an image 300 including red box on black background. The attended area 220 in image 300 is the red box. Red color is usually considered as bright color which easily attracts human attentions. However, the image 400 of Figure 4 cannot support this assumption. Rather, the block box of image 400 becomes the attended area 220 though red background occupies most of image 400. This phenomenon indicates that the color and size are not most pivotal factor for human perception, although human visual sensitivity has some intentions in color and size. Color contrast plays an important role in the human perception process.

[0035] Figures 5 and 6 show respective textured images 500 and 600, wherein oriented rectangles are surrounded by the texture. Figure 5 is illustrative of a weak texture area (central portions) surrounded by the strong texture patches around the border portions of the rectangle. Whereas in Fig. 6, a strong textured

area is surrounded by weak texture patches. Similar to color, the strength of texture does not greatly influence human perception, as does contrast. A similar conclusion can also be drawn from Figures 7 and 8. Additionally, the complexity of shape is not the main factor in human perception. From above comparisons, regions with high contrast are indicative of areas of rich information and are most likely to attract human attentions.

[0036] There are a number of known techniques to compute contrast, such as color contrast and luminance contrast. However, these techniques do not provide the type of contrast determinations needed for generation of the contrast-based saliency map 216 of Fig. 2. Rather, a more generic contrast is utilized. In particular, an effectual area is identified as one of perceiving stimulus, and is called a *perceive field*. The perceive field is the unit of contrast. The perceive field is analogous to a receptive field as identified by a human eye. An image with the size of *MxN* pixels is regarded as a *perceive field* with *MxN perception units*, if each *perception unit* contains one pixel. The contrast value $C_{i,j}$ on a *perception unit* (i, j)—a perceive field, is defined as follows:

$$C_{i,j} = \sum_{q \in \Theta} d\left(p_{i,j}, q\right) \tag{1},$$

where $p_{i,j}$ ($i \in [0, M]$, $j \in [0, N]$) and q denote the stimulus perceived by *perception units*, such as color. Θ is the neighborhood of *perception unit* (i, j). The size of Θ controls the sensitivity of *perceive field*. The smaller the size of Θ, the more sensitive the *perceive field* is. Parameter d is the difference between $p_{i,j}$ and q, which may employ any distance measure such as L_1 and L_2 distances.

Preprocessing to Resize, Transform, Quantize, and Divide Images

[0037] Preprocessing module 202 quantizes an original image 212 to generate quantized block image 214. Fig. 9 shows an example of an original image 212. Fig. 10 shows an example of the original image of Fig. 9, after it has been quantized. To this end, the preprocessing module 202 resizes the original image 212, while maintaining the original images aspect ratio. This effectively reduces computational complexity and maintains all images in a same configurable scale. If not already in a selected color space, such as LUV color space, the color space transformation is performed to transform the resized image to the selected color space. In this implementation, and since LUV space is consistent with human color perception, the resized image is transformed, for example, from RGB space to LUV space—a human perceptible color space.

[0038] The preprocessing module 202 color quantizes the transformed image. Human vision perception is more sensitive to the changes in smooth areas than to changes in areas of texture. To facilitate such perception, the color quantization operation makes color coarser in texture areas. In this implementation, well-known techniques to perform peer group filtering and perceptual color quantization of the transformed image are utilized for this operation. To further smooth texture areas and reduce computational cost, the preprocessing module 202 divides the quantized image into blocks, which for purposes of discussion are shown as quantized image blocks 214. Each quantized block is a perception unit with a perceive field of a certain number of pixels. In this implementation, the quantized image is divided into blocks of 8×8 pixels. LUV elements of each perception unit are computed separately.

Contrast Computation, Normalization, and Attended Data Extraction

[0039] At this point, preprocessing operations have completed, and the contrast computation and normalization (CCN) module 204 calculates a respective contrast for each of the quantized blocks 214. The CCN module 204 then smoothes and normalizes the calculated contrasts C_{ij} on the *perception units* to [0, 255]. This generates the contrast-based saliency map 216. Fig. 11 shows an example of a contrast-based saliency map 216 derived from the original image of Fig. 9 and the quantized image of Fig. 10. Three-levels of attention data, attended points 218, attended areas 220, and attended views are extracted from a contrast-based saliency map 216. Figs. 12-14 illustrate these contrast-based attention data that have been extracted by the CCN module 204 from the exemplary embodiment of the saliency map 216 of Fig. 11. In particular, Fig. 12 shows exemplary intermediate result of fuzzy partition 222, which have been generated by fuzzy growing as described below. Fig. 13 illustrates exemplary attended areas 220. Fig. 14 shows exemplary attended points 218.

[0040] In this implementation, colors in LUV space are used as stimulus on each perceive field, and the difference *d* is computed by Gaussian distance. Image attention analysis is performed on local contrast in the contrast-based saliency map 216, because this kind of saliency map not only reflects color contrast, but also reflects strength of texture. Additionally, areas close to the boundary of objects tend to have same or similar contrasts. Therefore, the contrast-based saliency map 216 further presents color, texture and approximate shape information, and thereby provides robust information for image attention analysis.

Attended Points Extraction

[0041] Attended point extraction module 206 directly detects and extracts attended points 218 from the contrast-based saliency map 216. Attended points 218 are points in the contrast-based saliency map 216 with local maximum contrast. Attended point detection is analogous to detection of a lowest level of human attention that has been directly caused by outside stimulus. As a result, attended points do not have any semantics. In this implementation, a maximum top five (5) points are extracted, because humans generally cannot focus on too many objects at the first glance. In a different implementation, some other number of attended points 218 is extracted from the contrast-based saliency map 216.

Attended Areas Extraction

[0042] Attended areas 220 are generated by the attended area extraction module 208. The result may be regarded as an extension of attended point detection. The operations include seed selection and "fuzzy growing". In this implementation, the contrast-based saliency map 216 is a gray-level image in which bright areas are considered to be attended areas, as shown in the example of Fig. 11. Use of a hard cut threshold is not effective for attended areas extraction, because gray-levels in saliency map show continuous variation, even with respect to a single object. Consequently, conventional region growing approaches based on one strict measure are not useful for this solution. Instead, fuzzy theory is employed, since it has been shown to be effective in imitating human mental behavior.

[0043] To extract attended areas 220, the contrast-based saliency map 216 is regarded as a fuzzy event modeled by a probability space. Contrast-based saliency

map 216 has L gray levels from g_0 to g_{L-1} and the histogram of saliency map is h_k, $k=0, ..., L-1$. Accordingly, the contrast-based saliency map 216 is modeled as a triplet (Ω, k, P), where $\Omega=\{g_0, g_1, ..., g_{L-1}\}$ and P is the probability measure of the occurrence of gray levels, i.e., $Pr\{g_k\} = h_k/\Sigma h_{k.}$. A membership function, $\mu_S(g_k)$, of a fuzzy set $S \in \Omega$ denotes the degree of certain properties, such as attended areas, unattended areas, and so on., possessed by gray level g_k. In fuzzy set notation, the membership function can be written as follows:

$$S = \sum_{g_k \in \Omega} \mu_S\left(g_k\right)/g_k \qquad (2).$$

[0044] The probability of this fuzzy event can be computed by

$$P(S) = \sum_{k=0}^{L-1} \mu_S\left(g_k\right)P_r\left(g_k\right) \qquad (3).$$

[0045] There are two classes of pixels in the contrast-based saliency map 216: attended areas and unattended areas of pixels. The two classes represent two fuzzy sets, denoted by B_A, and B_U, respectively, which are mutually exclusive. Thus, these two fuzzy sets partition the contrast-based saliency map 216 ("Ω"). In such a fuzzy partition, there is no sharp boundary between the two fuzzy sets, which is analogous to human perception mechanisms. Fuzzy c-partition entropy is utilized as a criterion to measure the fitness of a fuzzy partition. Theoretically, a fuzzy c-partition is determined by $2(c-1)$ parameters. Thus, it is useful to find substantially the best combinations of these parameters, which is considered to be a combinatorial optimization problem. Simulated annealing or genetic algorithms are generally used to solve this type of optimization problem. These are very processing and time intensive operations. However, only two (2) parameters are

used in the present algorithm due to 2-partition. Therefore, this implementation of the attended area extraction module 208 utilizes an exhaust search to find optimal result without involving high computational complexity.

[0046] In the saliency map Ω, considering the two fuzzy events, attended areas B_A, and unattended areas B_U, the membership functions of fuzzy events are defined in (4) and (5), respectively as follows:

$$\mu_A = \begin{cases} 1 & x \geq a \\ \dfrac{x-u}{a-u} & u < x < a \\ 0 & x \leq u \end{cases} \qquad (4);$$

$$\mu_U = \begin{cases} 0 & x \geq a \\ \dfrac{x-a}{u-a} & u < x < a \\ 1 & x \leq u \end{cases} \qquad (5);$$

wherein x is an independent variable denoting gray level, and a and u are parameters determining shape of the above two membership functions. If an optimization objective function is satisfied, the optimal parameters a and u are obtained. Gray-levels greater than a have the membership of 1.0 for fuzzy set B_A, which means the pixels with these gray-levels definitely belong to the attended areas. In distinction, when the gray levels is smaller than u, the membership for fuzzy set B_A becomes zero (0), which means the pixels with these gray-levels do not belong to the attended areas. Similarly, B_U has opposite variation form. While, the pixels with the gray-levels between a and u have the membership of (0, 1) for fuzzy sets B_A and B_U according to the definition (4) and (5), respectively.

[0047] Assuming that the prior probabilities of the attended areas 220 and the unattended areas are approximately equal, the optimal partition entails that the difference between the prior entropies of attended areas and that of unattended

areas reaches the minimum. A minimal difference of entropy as a metric to obtain optimal threshold for image segmentation is modified in view of a fuzzy set definition as follows:

$$\Gamma(a,u) = \left[H_A(a,u) - H_U(a,u) \right]^2 \qquad (6),$$

wherein, $H_A(a, u)$ and $H_U(a, u)$ are prior entropies of fuzzy sets, attended areas 220 and unattended areas (e.g., see the unattended area(s) of "other data" 224 of Fig. 2), respectively. They are calculated as:

$$H_A(a,u) = -\sum_{k=0}^{L-1} \frac{\Pr(g_k)}{P(B_A)} \ln \frac{\Pr(g_k)}{P(B_A)} \qquad (7);$$

$$H_U(a,u) = -\sum_{k=0}^{L-1} \frac{\Pr(g_k)}{P(B_U)} \ln \frac{\Pr(g_k)}{P(B_U)} \qquad (8);$$

wherein $P(B_A) = \sum_{k=0}^{L-1} \mu_A \Pr(g_k)$ and $P(B_U) = \sum_{k=0}^{L-1} \mu_U \Pr(g_k)$ according to equation (3).

[0048] Global minima of $\Gamma(a, u)$ indicates the optimal fuzzy partition, i.e., optimal parameters a and u are found. This criterion can be expressed as:

$$(a, u) = \mathbf{arg\ min}\ (\ \Gamma(a, u)\ ) \qquad (9).$$

[0049] With the optimal a and u, fuzzy growing is performed on the contrast-based saliency map 216. A number of initial attention seeds are utilized. Exemplary criteria for seed selection include, for example: the seeds have maximum local contrast; and the seeds belong to the attended areas 220. Sequentially, starting from each seed, the pixels with the gray-levels satisfying the following criteria are grouped by the attended area extraction module 208 as follows:

$$C_{i,j} \le C_{seed} \text{ and } C_{i,j} > s \qquad (10),$$

wherein, $s = (a + u)/2$. In this implementation, the probabilities of gray-level s belong to attended areas 220 and unattended areas are all 0.5, see equations (4)

and (5). Then, the new group members are used as seeds for iterative growth. Such fuzzy growing process simulates a bottom-up search process in human perception.

[0050] Fig. 12 illustrates an exemplary fuzzy 2-partition of the contrast-based saliency map 216 with three layers, which denote a gray level higher than a (highest), s (middle), and u (lowest), correspondingly. Fig. 13 shows exemplary results of fuzzy growing, two main objects in scene being accurately detected and segmented.

[0051] In view of the foregoing, attended area seeds are the subset of attended points 218. Points 218 are selected for seeds if they have contrasts greater than a. Then, from each seed, fuzzy growing is carried out until no candidate of perception units can be grouped. This process simulates early stage of human perception during which human search a semantic object looks like what has already been seen.

Attended View Extraction

[0052] Attended view extraction module 210, formulated in view of a non-computational Gestalt law of psychology of visual form, extracts an attention center 224 as well as an attended view 222 from the saliency map 216. In particular, it is assumed that since visual forms may possess one or several centers of gravity about which the form is organized, that there is a center of gravity (i.e., an attention center 224) in a saliency map 216, which corresponds to the vision center of the original image 212. Based on the attention center 224, the whole image is organized for information maximization.

[0053] In this implementation, an attended view 222 is a rectangle $V(C, W, H)$, where C denotes attention center, W and H are the width and height of

rectangle respectively. If contrast (gray level) in a saliency map 216 is regarded as density, the attention center 224 is the centroid of the saliency map 216. Similarly, there is a relationship between the size of attended view 222 and the 1st order central moment of the saliency map. Specifically, let (x_0, y_0) denote attention center, and (w', h') denote the 1st order central moment of saliency map, the attention center and the attended view's width and height are computed by (11) and (12) respectively,

$$\begin{cases} x_0 = \dfrac{1}{CM} \sum_{j=0}^{N-1} C_{i,j} \times i \\ \\ y_0 = \dfrac{1}{CM} \sum_{i=0}^{M-1} C_{i,j} \times j \end{cases} \tag{11},$$

where $CM = \displaystyle\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} C_{i,j}$ is the 0th order moment of saliency map.

$$\begin{cases} W = 2w = 2\alpha \cdot w' \\ H = 2h = 2\alpha \cdot h' \end{cases} \tag{12},$$

where $\alpha > 1$ is a constant coefficient. Parameters w' and h' are computed by the 1st order central moments of saliency map 216 along x-axis and y-axis respectively, and the 0th order moment CM, expressed by (13).

$$\begin{cases} w' = \dfrac{1}{CM} \sum_{j=0}^{N-1} C_{i,j} \times \|i - x_0\| \\ \\ h' = \dfrac{1}{CM} \sum_{i=0}^{M-1} C_{i,j} \times \|j - y_0\| \end{cases} \tag{13}.$$

[0054] The operation of attended view extraction can be viewed as the last stage of human perception. That is, when a human completes attention searching,

views are typically adjusted as a function of the image attention center and attention distributions in the image as a whole.

[0055] Referring to Figs. 15 through 44, all of which are shown on page 5 of the drawings, the Figures in column 1 illustrate examples of original images 212, the Figures in column 2 show examples of contrast-based saliency maps 216, the Figures of column 3 illustrate examples of attended views 222, the Figures of column 4 show exemplary attended areas 220, and the Figures of column 5 illustrate examples of attended points 218. All images of a row 1 though 6 of the Figures 15-44 are illustrative of results derived from the systems and methods described herein applied to the leftmost Figure in the row.

An Exemplary Procedure

[0056] Fig 45 shows an exemplary procedure 4500 for generating a contrast-based saliency map for image attention analysis. The operations of the procedure are implemented and described with respect to program modules of Fig. 2. (The left-most digit of a component reference number identifies the particular figure in which the component first appears). At block 4502, the preprocessing module 202 preprocesses an original image 212. Such preprocessing operations include, for example, image resizing, color transformation, and quantization operations, resulting in quantized image blocks 214. At block 4504, the contrast computation and normalization module 204 generates a contrast-based saliency map 216 from the quantized image blocks 214. At block 4506, attended point extraction module 206 extracts attended points 218 from the contrast-based saliency map 216. At block 4508,

attended area extraction module 206 extracts attended area 220 from the contrast-based saliency map 216 in view of the attended points 218. At block 4510, the attended view extraction module 210 extracts attended view 222 from the contrast-based saliency map 216.

Conclusion

[0057] The described systems and methods for generating a contrast-based saliency map for image attention analysis. Although the systems and methods have been described in language specific to structural features and methodological operations, the subject matter as defined in the appended claims are not necessarily limited to the specific features or operations described. Rather, the specific features and operations are disclosed as exemplary forms of implementing the claimed subject matter. For instance, with the provided three-level image attention analysis, performance of visual perception systems, multimedia systems, and information searching in large image library can be greatly improved in accuracy, speed and display aspects. Additionally, integration with other image analysis applications, such as a face detection application, can provide additional information to modify attended areas 220 and attended view 222 results (e.g., in view of face rectangle semantics). Analogously, attended areas 220 and attended view 222 results can be used to speed up the process of other image analysis applications such as one that implements a face searching algorithm.